SPATIAL AND ANGULAR RECONSTRUCTION OF LIGHT FIELD BASED ON DEEP GENERATIVE NETWORKS

Nan Meng *Tianjiao Zeng* *Edmund Y. Lam*

Department of Electrical and Electronic Engineering, University of Hong Kong, Pokfulam, Hong Kong

ABSTRACT

Light field (LF) cameras often have significant limitations in spatial and angular resolutions due to their design. Many techniques that attempt to reconstruct LF images at a higher resolution only consider either spatial or angular resolution, but not both. We propose a generative network using high-dimensional convolution to improve both aspects. Our experimental results on both synthetic and real-world data demonstrate that the proposed model outperforms existing state-of-the-art methods in terms of both peak signal-to-noise ratio (PSNR) and visual quality. The proposed method can also generate more realistic spatial details with better fidelity.

***Index Terms*—** Light field reconstruction, generative adversarial networks, computational imaging, high-dimensional convolution, deep learning

1. INTRODUCTION

Light field (LF) imaging typically makes use of a compact camera array to capture multiple images of the scene from slightly different viewpoints [1]. Compared with traditional 2D images, a light field also records the angular direction information of the light rays, which enables additional applications, such as depth estimation [2], 3D modeling [3], digital refocusing [4], and virtual reality. However, the benefits from such additional angular information are usually obtained at the expense of the spatial resolution, which should therefore be augmented by computational methods [5, 6]. A common approach for light field super-resolution (LFSR) is to consider it as two parts, namely spatial super-resolution (SR) and view synthesis, and solve them separately [7, 8]. However, the LF data have a particular structure that should be preserved when enhancing the resolution, leading to the need for better algorithmic approaches that also model the light distribution.

Many existing algorithms are based on disparity information for LFSR, but they usually exhibit noticeable drawbacks. For example, Wanner and Goldluecke [9] first exploited the slopes in epipolar plane images (EPIs) to calculate the disparity map for every individual low-resolution (LR) view. The obtained depth is then used to super-resolve the LFs in a variational optimization framework. Since the disparity calculation remains challenging at low spatial resolution, such method easily results in significant artifacts in the textured and occlusion regions. Mitra and Veeraraghavan [7] proposed a patch-based model based on Gaussian mixture model and reconstructed the patches according to the disparity value. However, assigning a constant disparity to each sampled LF patch results in severe artifacts at depth discontinuities in the reconstructed views. Apart from improving spatial resolution, some have studied approaches to synthesize novel views. Pearson et al. [10] introduced an automatic depth layer-based method for synthesizing an arbitrary view from a set of existing views. They rendered the view using a probabilistic interpolation based on depth information on a small set of sub-aperture images. However, these methods rely heavily on the estimated depth, which is sensitive to occluded regions and reflective surfaces.

Recently, a few methods based on convolutional neural network (CNN) were proposed [11, 12]. Yoon et al. [13] considered the cascade of two CNNs to super-resolve the given views. However, the design of their model underusee the potential of the entire angular information, seriously limiting the performance. Meanwhile, Kalantari et al. [8] used two sequential CNNs to model disparity and estimate color simultaneously. However, since the framework was depth-dependent and training process required fixed sampling patterns, the approach resulted in artifacts in occluded and reflective regions. Wu et al. [14] exploited the texture structure of the EPI and modeled the view synthesis as the angular restoration of the EPI. Their EPI-based method got rid of depth estimation, but severely limited the accessible information of the model, resulting the artifacts in the generated novel views.

Given the inherent geometry of LF data, reconstruction algorithms should involve information from both spatial and angular dimensions [15]. We therefore apply the high-dimensional convolution (HDC) to process LF data and wrap the operation into a layer to construct the deep network. Inspired by [16, 17], the generative adversarial networks (GANs) are powerful in generating plausible natural images with high perceptual quality. Therefore, we propose a generative model named LFGAN, which incorporates the HDC layers into a GAN framework to learn the high correlations among neighboring LF views. We also define a novel loss function to encourage LFGAN in recovering more realistic spatial details. Experimental results demonstrate that our LF-

GAN can enhance realistic spatial details and generate novel view with good fidelity.

2. METHOD

2.1. Problem Formulation

The LF reconstruction deals with the recovery of the high-resolution (HR) data $I^H(x, y, u, v) \in \mathbb{R}^{\gamma_s H \times \gamma_s W \times \gamma_a S \times \gamma_a T}$ from the corresponding LR data $I^L(x, y, u, v) \in \mathbb{R}^{H \times W \times S \times T}$, with the spatial and angular super-resolution factors γ_s and $\gamma_a \in \mathbb{Z}^+$. We cast the reconstruction task into the tensor restoration problem which can be described as

$$I^S(x, y, u, v) = g(I^L(x, y, u, v), \Theta), \tag{1}$$

where $\Theta = \{\theta^{(0)}, \cdots, \theta^{(K-1)}\}$ represents the parameters of the networks and I^S stands for the super-resolved LF. The function $g(\cdot)$ describes the learned mapping from LR to HR light field images, where in this paper it is formulated as an adversarial generative network. All model parameters are optimized to reduce the loss $\ell(\cdot)$, which measures the difference between I^S and I^H. Therefore, the restoration task is

$$\Theta^* = \arg\min_{\Theta} \ \ell\big(I^H, g(I^L; \Theta)\big). \tag{2}$$

However, different from single-image super-resolution but in a way more similar to holographic reconstruction [18, 19], LFSR requires the algorithm $g(\cdot)$ to possess the ability to recover high-frequency spatial details while preserving angular correlations. Given the complex structural relations among different LF coordinates, we use the HDC operation instead of traditional 2D convolution in our network. Furthermore, we also carefully design the loss function to encourage the network to reconstruct realistic spatial details.

2.2. High-dimensional Convolution

The HDC operation is used to enable our network to enforce the LF coordinates relations. Each input LF captures the geometry information of the scene, which is wrapped and stored in its structural high-dimensional data. The intrinsic limitation of 2D (or 3D) convolution makes it hard to handle LF problem, and hence existing learning-based methods simplify the reconstruction to only model the spatial-angular relation (EPI images) [14] or the angular correlations (sub-aperture image sequence) [8, 13, 20]. On the contrary, in our proposed network, each convolutional layer fully exploits the information of all coordinates by applying HDC. Stacking many such layers leads to filters that become increasingly global and therefore the network can use more context to predict the spatial details. Meanwhile, the convolution along angular coordinates preserves geometry information. As the GAN-based model maintains the ability to drive the reconstruction towards HR image manifold, we incorporate the HDC with GANs framework and construct both generator G and discriminator D using HDC layer.

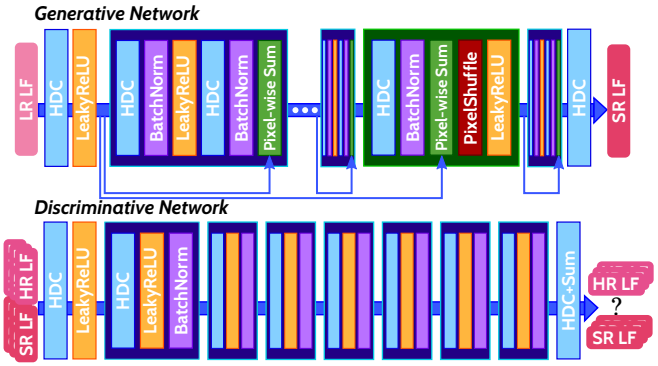

Fig. 1. The proposed LFGAN model by incorporating the HDC layer into the GANs framework.

Fig. 1 presents the entire framework of our proposed LFGAN. Different types of layers are denoted by different colors. The training process can be considered as solving the min-max problem [17]

$$\begin{aligned}\min_{\theta_G} \max_{\theta_D} \quad & \mathbb{E}_{I^H \sim \pi(I^H)} \log D\big(I^H\big) + \\ & \mathbb{E}_{I^L \sim \pi(I^L)} \log\big[1 - D\big(G(I^L)\big)\big]\,,\end{aligned} \tag{3}$$

where θ_G and θ_D stand for the parameters of generator and discriminator, respectively.

2.3. Loss Function

To efficiently drive our proposed generator learning the light distribution, we define a novel spatial-angular loss function. Such loss is formulated as the weighted sum of a spatial content loss ℓ_S, an angular loss ℓ_A based on mean square error (MSE), and an adversarial loss component ℓ_G, such that

$$\ell = \alpha \cdot \ell_S + \beta \cdot \ell_A + \gamma \cdot \ell_G. \tag{4}$$

The scalars α, β and γ denote the weights of each loss. The spatial loss ℓ_S is calculated based on the features of ReLU activation layers of a pre-trained 19 layer VGG network. We compute such spatial loss on each sub-aperture image of LF, and therefore

$$\ell_S = \frac{1}{ST} \sum_{s=1}^{S} \sum_{t=1}^{T} \Big[f(I^H)_{s,t} - f\big(G(I^L)\big)_{s,t} \Big]^2, \tag{5}$$

where $f(\cdot)$ indicates the feature map described in [20].

The angular loss ℓ_A is defined based on EPIs which contain the angular information. We use the MSE between the original and super-resolved EPI features to encourage our generator to maintain the angular correlations while enhancing the spatial resolution, i.e.,

$$\ell_A = \sum_{y=1}^{Y} \sum_{t=1}^{T} \Big(E^{I^H}_{y,t}(x, s) - E^{I^S}_{y,t}(x, s) \Big)\,. \tag{6}$$

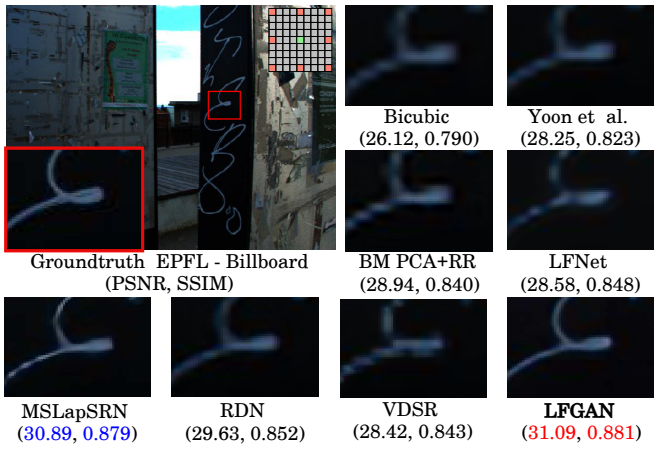

Fig. 2. Visual and quantitative (PSNR, SSIM) comparison of the proposed LFGAN against other SR methods on the center view of a real-world scene: EPFL–Billboards (for $4\times$ SR).

The third component is the adversarial generative loss, i.e.,

$$\ell_G = \sum_{n=1}^{N} \log\Big[1 - D\Big(G\big(I^L(x,y,s,t)\big)\Big)\Big]. \qquad (7)$$

3. EXPERIMENTS AND RESULTS

We compare the proposed framework with some state-of-the-art methods on several aspects. For spatial SR, they include some recent LF reconstruction, such as Yoon et al. [13], BM PCA+RR [20], LFNet [11], and single image super-resolution (SISR) methods, including MSLapSRN [21], RDN [22], and VDSR [23]. These are applied on every sub-aperture image of a LF. For angular SR, we compare the approaches proposed by Wu et al. [14] and Kalantari et al [8].

3.1. Experimental Settings

We implement the HDC operation and subsequently the LFGAN using Tensorflow. Our training data are sampled from Lytro Archive [24] (not including the scenes in reflective and occlusion categories) and Fraunhofer densely-sampled high-resolution dataset [25]. To obtain spatial LR data, the training data are downsampled according to the imaging model

$$I^L = \delta(B * I^H) + \eta, \qquad (8)$$

where η represents the additive noise, $\delta(\cdot)$ is the nearest neighbor downsampling operator and B is the Gaussian kernel. To get LR angular data, we sparsely sampled the views of LF data by a factor γ_a. The learning rate is initialized to 10^{-5} and decreased by a factor of 0.1 for every 10 steps.

	Buddha	Mona	Reflectives 20	Occlusions 20
Bicubic	28.58	29.44	31.19	28.52
Yoon et al.	29.84	31.40	31.42	28.86
BM PCA+RR	30.43	32.68	33.07	30.45
LFNet	30.93	32.47	33.85	30.37
VDSR	30.48	31.39	32.32	29.84
MSLapSRN	30.98	32.74	32.43	30.85
RDN	30.99	31.80	33.86	31.46
LFGAN	**31.93**	**32.97**	**35.24**	**33.15**

Table 1. Quantitative evaluation (PSNR) on synthetic LF and real-world LF for $\gamma_s = 4$. All numbers are measured in dB.

3.2. Spatial Super-resolution

The spatial SR enhances the resolution of every sub-aperture image of LF by a factor of γ_a. Our proposed model fully exploits the spatio-angular information by employing the HDC layer. As a consequence, the LFGAN is able to provide superior visual results on the spatial details. Fig. 2 compares the visual results of LFGAN against several advanced methods. The average peak signal-to-noise ratio (PSNR) is reported under every reconstructed close-up image.

Quantitative performances of our proposed model in HCI synthetic scenes [26] and Lytro Archive [24] real-world scenes for $4\times$ spatial reconstruction are presented in Table 1. We evaluate our methods against other state-of-the-art methods as well as the standard interpolation method (Bicubic) as the baseline. For real-world evaluation, we randomly select 20 scenes from two challenging categories — reflective and occlusion. As seen in the Table, LFGAN outperforms all other reference methods on the mean PSNR measurement.

3.3. Angular Super-resolution

The angular SR, also named view synthesis, aims to reconstruct the novel views based on sparsely sampled input views. We compare LFGAN with two current learning-based approaches, namely Kalantari et al. [8] and Wu et al. [14]. Fig. 3 presents the visual results of LFGAN against these two methods. The former is designed based on depth information. Therefore, it tends to fail on the region with complex occlusions (the "fence" region) and result in ghosting artifacts (the "wire" region). The latter generates novel view by restoring angular information on the EPI. Therefore, the structural information of LF is not fully exploited during the reconstruction. In some regions with fine texture (e.g., the "feelers" of papillon), such method easily results in artifacts. Compared with these methods, the HDC layer enables LFGAN to use all the information in every dimension of LF, leading to a more robust performance even in challenging scenes.

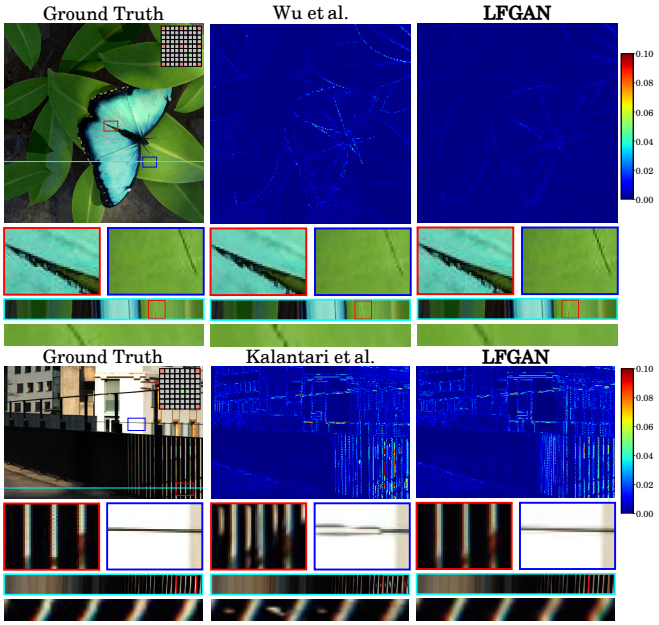

Fig. 3. Comparison of the proposed LFGAN against recent state-of-the-art synthesis methods on the synthetic scene (Top: *HCI–Papillon*) and the real-world scene (Bottom: *EPFL–Black Fence*).

3.4. Spatial-angular Super-resolution

The spatial-angular SR both reconstructs the spatial resolution of each sub-aperture image and also generates a dense LF based on the sparsely sampled LF. The ability to calculate HD data directly allows our proposed LFGAN to recover both spatial and angular information simultaneously. We test the model performance for $2\times$ spatial resolution enhancement while at the same time generating 9×9 views from 5×5 views. The visual results are presented in Fig 4, in comparison with Yoon et al. [13] and LFSR [27], which can also restore both spatial and angular resolution. Compared with the performance of our method, the results of Yoon et al. tend to be blurred while LFSR leads to artifacts near the edges.

4. CONCLUSION

We have described a generative framework employing HDC layer for LF spatial and angular information reconstruction. The proposed LFGAN benefits from training on a novel loss, which drives the generator to reconstruct high-quality LF with realistic spatial details. We compared the performance of our model against state-of-the-art methods for spatial, angular and spatio-angular SR tasks separately. The experimental results demonstrate that our LFGAN has the capacity to reconstruct spatial details with good fidelity and with high quantitative score on PSNR.

This work is supported in part by the Hong Kong Research Grant Council (17203217 and 17201818) and the University of Hong Kong (104005009, 104005438).

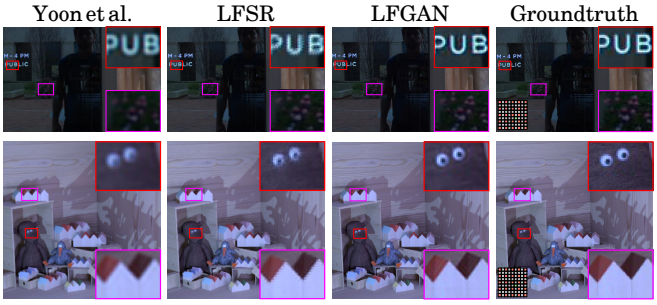

Fig. 4. Visual comparison for spatial-angular reconstruction. Both spatial and angular resolutions have been downsampled with the spatial factor $\gamma_s = 2$ and the angular factor $\gamma_a = 2$ (generate 9×9 views from 5×5 views).

5. REFERENCES

[1] Edmund Y. Lam, "Computational photography with plenoptic camera and light field capture: Tutorial," *Journal of the Optical Society of America A*, vol. 32, no. 11, pp. 2021–2032, November 2015.

[2] Xing Sun, Zhimin Xu, Nan Meng, Edmund Y. Lam, and Hayden K.-H. So, "Data-driven light field depth estimation using deep convolutional neural networks," in *IEEE International Joint Conference on Neural Networks*, July 2016, pp. 367–374.

[3] Robert Prevedel, Young-Gyu Yoon, Maximilian Hoffmann, Nikita Pak, Gordon Wetzstein, Saul Kato, Tina Schrödel, Ramesh Raskar, Manuel Zimmer, Edward S Boyden, and Alipasha Vaziri, "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy," *Nature Methods*, vol. 11, no. 7, pp. 727–730, 2014.

[4] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report*, vol. 2, no. 11, pp. 1–11, 2005.

[5] Zhimin Xu and Edmund Y. Lam, "Light field superresolution reconstruction in computational photography," in *OSA Topical Meeting in Signal Recovery and Synthesis*, July 2011, p. SMB3.

[6] Zhimi Xu and Edmund Y. Lam, "A spatial projection analysis of light field capture," in *OSA Frontiers in Optics*, October 2010, p. FWH2.

[7] Kaushik Mitra and Ashok Veeraraghavan, "Light field denoising, light field superresolution and stereo camera

based refocussing using a GMM light field patch prior," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, July 2012, pp. 22–28.

[8] Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Transactions on Graphics*, vol. 35, no. 6, pp. 193:1–193:10, November 2016.

[9] Sven Wanner and Bastian Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, March 2014.

[10] James Pearson, Mike Brookes, and Pier Luigi Dragotti, "Plenoptic layer-based modeling for image based rendering," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3405–3419, June 2013.

[11] Yunlong Wang, Fei Liu, Kunbo Zhang, Guangqi Hou, Zhenan Sun, and Tieniu Tan, "LFNet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4274–4286, 2018.

[12] Nan Meng, Hayden K.-H. So, Xing Sun, and Edmund Y. Lam, "High-dimensional dense residual convolutional neural network for light field reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, submitted.

[13] Youngjin Yoon, Hae-Gon Jeon, Donggeun Yoo, Joon-Young Lee, and In So Kweon, "Light-field image super-resolution using convolutional neural network," *IEEE Signal Processing Letters*, vol. 24, no. 6, pp. 848–852, 2017.

[14] Gaochang Wu, Mandan Zhao, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu, "Light field reconstruction using deep convolutional network on EPI," in *IEEE Conference on Computer Vision and Pattern Recognition*, November 2017, pp. 1638–1646.

[15] Nan Meng, Xing Sun, Hayden K.-H. So, and Edmund Y. Lam, "Computational light field generation using deep nonparametric Bayesian learning," *IEEE Access*, vol. 7, pp. 24990–25000, February 2019.

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *The 27th Neural Information Processing Systems Advances*, pp. 2672–2680. December 2014.

[17] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 105–114, November 2017.

[18] Zhenbo Ren, Zhimin Xu, and Edmund Y. Lam, "End-to-end deep learning framework for digital holographic reconstruction," *Advanced Photonics*, vol. 1, no. 1, pp. 016004, January 2019.

[19] Zhenbo Ren, Zhimin Xu, and Edmund Y. Lam, "Learning-based nonparametric autofocusing for digital holography," *Optica*, vol. 5, no. 4, pp. 337–344, April 2018.

[20] Reuben A Farrugia, Christian Galea, and Christine Guillemot, "Super resolution of light field images using linear subspace projection of patch-volumes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1058–1071, August 2017.

[21] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," *arXiv preprint arXiv:1710.01992*, 2017.

[22] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu, "Residual dense network for image super-resolution," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2472–2481.

[23] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *International Conference on Learning Representations*, 2015.

[24] "Stanford lytro light field archive," http://lightfields.stanford.edu/.

[25] Matthias Ziegler, Ron op het Veld, Joachim Keinert, and Frederik Zilly, "Acquisition system for dense light-field of large scenes," in *IEEE Conference on The True Vision-Capture, Transmission and Display of 3D Video*, February 2017, pp. 1–4.

[26] Sven Wanner, Stephan Meister, and Bastian Goldluecke, "Datasets and benchmarks for densely sampled 4D light fields.," in *Vision, Modeling, and Visualization*. Citeseer, 2013, pp. 225–226.

[27] M. S. K. Gul and B. K. Gunturk, "Spatial and angular resolution enhancement of light fields using convolutional neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2146–2159, May 2018.